The Glass Cannon Phenotype: High Predicted Benefit but Low Robustness to Biological and Dosage Perturbations

Per Magnus Swedenborg[1,]

[1] DNAI Biotech Correspondence: per.swedenborg@dnai.bio | ORCID: 0009-0008-2750-9735

Keywords: Precision Oncology, Treatment Fragility, Digital Twins, Decision Support, Risk Stratification

---

Abstract

Purpose: Precision oncology algorithms typically rank treatments by maximizing expected benefit (efficacy). However, these rankings often ignore robustnesswhether the recommendation remains consistent when dose or key biological parameters vary slightly within plausible ranges. We identify a Glass Cannon phenotype: patients for whom models predict high therapeutic benefit, but whose treatment plans are highly sensitive to perturbations.

Methods: We analyzed 9,393 patients across 33 cancer types from The Cancer Genome Atlas (TCGA) using DNAI, a physics-constrained cancer digital twin platform. We computed two signals for each patient: (1) a **Benefit Score**, defined as the difference between predicted 5-year survival under the top-ranked treatment class versus no treatment; and (2) a composite **Fragility Score (F)**, derived from four *in silico* sensitivity analyses (dose sensitivity, biological parameter noise, treatment substitution variance, and model uncertainty). Patients were stratified into quadrants based on Benefit (Median split) and Fragility (Top Tercile split).

Results: We identified a distinct Glass Cannon cohort (7.0%, $N = 653$) characterized by high predicted benefit but high fragility ($F > 66^{th}$ percentile). Glass Cannon patients experienced significantly worse Overall Survival (median 478 days, 71.1% event rate) than Stable Non-Responders (median 699 days, 29.2% event rate), despite having higher model-estimated efficacy. The phenotype was most prevalent in immune-active tumors, specifically Kidney Renal Clear Cell Carcinoma (KIRC, 41.9%) and Skin Cutaneous Melanoma (SKCM, 23.5%). Fragility independently predicts early treatment failure (AUC = 0.788, $P < 10^{241}$).

Conclusion: High predicted benefit does not imply treatment safety. The Glass Cannon phenotype represents a critical failure mode where aggressive treatment plans are theoretically optimal but operationally brittle. We propose a decision checklist to distinguish between mechanistic volatility (requiring dose protection) and model uncertainty (requiring abstention). This represents a risk-stratification signal, not a causal estimate of treatment effect.

Introduction

The central tenet of precision oncology is the identification of the right drug for the right patient by maximizing expected treatment benefit. Computational decision support systems (CDSS) and molecular tumor boards typically rank therapeutic options based on predicted efficacywhether derived from genomic biomarkers, transcriptomic signatures, or machine learning models. The

standard assumption is that the treatment with the highest predicted magnitude of effect is the optimal clinical choice.

However, clinical decisions require not only expected benefit, but also robustness to routine deviations in care and biology. Clinicians frequently encounter patients with similar biomarker profiles where one responds durably, while the other progresses rapidly following a minor dose reduction, a delay in administration, or a slight shift in tumor microenvironment status. Current algorithmic rankings cannot distinguish between these two scenarios. In engineering, a system that performs well under narrow conditions but fails under slight perturbations is termed fragile.

We posit that patients with high predicted benefit but high plan fragilitya phenotype we term the Glass Cannonrepresent a high-risk subgroup currently misclassified by standard precision medicine algorithms. For example, a patient might be predicted to have a profound response to immunotherapy, but the model indicates this response collapses if the immune infiltration parameter varies by even 5%. Standard ranking metrics would flag this patient as a top candidate; a robustness-aware system would flag them as high-risk.

To quantify this phenomenon, we utilized the DNAI platform to perform simulation-based sensitivity analyses on patient-specific treatment plans. Unlike static risk scores, this approach simulates counterfactual tumor trajectories under varying conditions of dosage, biological noise, and model uncertainty. In this study of 9,393 patients from The Cancer Genome Atlas (TCGA), we demonstrate that Treatment Fragility is orthogonal to Predicted Benefit, identify the Glass Cannon subgroup, and propose a checklist to integrate robustness into clinical decision-making.

Methods

Overview of Computed Signals

For every patient, we computed two distinct metrics. First, the **Benefit Score** estimates the magnitude of survival gain from the optimal treatment class compared to no treatment. Second, the **Fragility Score** measures the variance of that survival prediction when the simulation is subjected to noise. The intersection of these two metrics defines our quadrants of interest.

Data Source & Processing

We analyzed 9,393 patients across 33 cancer types from The Cancer Genome Atlas (TCGA) Pan-Cancer Atlas [1]. Multi-omics data (RNA-seq, Copy Number Variation, Mutation, Methylation) were processed to produce a 328-dimensional compressed molecular profile (latent representation, z) using the DNAI VAE v5.10. This representation encodes proliferation, pathway activation (50 MSigDB Hallmarks [2]), and tumor microenvironment (TME) context.

The DNAI Model (Specialist Path A)

We utilized the DNAI Specialist Model (v3.2), a hypernetwork architecture that maps patient molecular profiles and histopathology embeddings (derived from UNI2-h [3]) to the parameters of a governing Neural Ordinary Differential Equation (ODE) [4]. The ODE models tumor burden (N) over time under therapy (D) and immune pressure (I):

$$\frac{dN}{dt} = N\,(1\ \frac{N}{K})\ \ {}_d{}^d D_d(t) N\ \ I(t) N$$

where is the tumor growth rate, $_d$ is drug sensitivity, and is immune kill efficiency. Parameters are constrained to biologically plausible ranges ($[0, 0.3]$, $[0, 1]$, > 0).

Counterfactual Ranking Estimation (Benefit)

We quantify predicted benefit by comparing simulated survival under different treatment classes. The model predicts outcomes conditional on a learnable treatment embedding representing 6 coarse modalities: Chemotherapy, Immunotherapy, Targeted Therapy, Hormone Therapy, Radiation, and None. * **Ranking Score:** For each patient, we simulate 5-year survival probability $S(5yr|x, t)$ under all treatment conditions.

$$\text{Benefit Score} = S(5yr|x, t_{best})\ \ S(5yr|x, t_{none})$$

Quantifying Treatment Fragility

Fragility measures how much the predicted survival changes when we perturb dose, parameters, treatment choice, and model stochasticity. For a given patient x and their top-ranked treatment t_{best}, we perform four *in silico* sensitivity analyses:

1. **Dose Sensitivity (V_{dose}):** We apply 11 scalar multipliers m $\{0.5, 0.6, \ , 1.5\}$ (linearly spaced) to the dosage function $D(t)$ in the ODE solver. We compute the variance of the resulting 5-year survival probabilities.
2. **Parameter Noise (V_{param}):** We perturb the predicted ODE parameters $(, ,)$ using the models Bayesian heads (sampling from predicted $,)$. We simulate 200 trajectories and compute the variance of $S(5yr)$.
3. **Treatment Substitution (V_{subst}):** We compute the variance of $S(5yr)$ across all 6 treatment classes (Chemotherapy, Immunotherapy, Targeted Therapy, Hormone Therapy, Radiation, and None). High variance implies the benefit is highly specific to one modality (regimen-dependent).
4. **Model Uncertainty (V_{unc}):** We perform 20 forward passes using Monte Carlo Dropout [5] (dropout layers active, BatchNorm statistics frozen). We compute the variance of the predicted survival probabilities.

The Composite Score:

$$F = 0.30\ \ R(V_{dose}) + 0.30\ \ R(V_{param}) + 0.25\ \ R(V_{subst}) + 0.15\ \ R(V_{unc})$$

Where $R()$ denotes the global percentile rank (01).

Quadrant Definition & Thresholds

To avoid arbitrary thresholds, we utilized population quantiles for stratification. We label the high-benefit/high-fragility quadrant Glass Cannon: * **High Benefit:** Score > Population Median (50^{th} percentile). * **High Fragility:** Score > Top Tercile (66^{th} percentile). * **Out-of-Distribution (OOD) Score (S_{OOD}):** Defined as the Mahalanobis distance [6] of the patients molecular profile z from the training centroid. High S_{OOD} (> 95th percentile) indicates low transportability.

Statistical Analysis

The model was trained on an 85/15 random split (seed=42) with 7,985 training and 1,408 validation patients. Benefit and fragility scores were computed using the trained model on the full cohort (N = 9,393). Survival comparisons used Kaplan-Meier estimation [7] and Mann-Whitney U tests. Predictive performance was assessed via Concordance Index (C-index) [8].

Experiments & Results

1. The Four Quadrants of Response

For each patient, we computed a predicted Benefit Score and a Fragility Score. We then stratified patients using a median split for benefit and a top-tercile split for fragility to form four quadrants (Table 1). The Glass Cannon phenotype (Quadrant 2) comprises 7.0% of the cohort.

Table 1: Survival Outcomes by Benefit Œ Fragility Quadrant (N = 9,393)

Quadrant	Definition	N (%)	Median OS (Days)	Event Rate	C-index
Q1: Solid Responder	High Ben / Low Frag	4,043 (43.0%)	830	19.5%	0.838
Q2: Glass Cannon	High Ben / High Frag	653 (7.0%)	478	71.1%	0.769
Q3: Stable Non-Responder	Low Ben / Low Frag	4,154 (44.2%)	699	29.2%	0.834
Q4: Fragile Non-Responder	Low Ben / High Frag	543 (5.8%)	350	70.7%	0.781

Median OS: Unadjusted, pooled across 33 cancer types. C-index: within-quadrant concordance of the model risk score. Q1 vs Q2: $P = 1.69 \text{ Œ } 10^{31}$*; Q3 vs Q4:* $P = 3.70 \text{ Œ } 10^{36}$.

2. The Glass Cannon Survival Paradox

The most striking finding is the survival inversion between Q2 and Q3. Glass Cannon patients have higher predicted benefit than Stable Non-Responders, yet their survival outcomes are significantly worse.

Figure 1: Benefit vs. Fragility Scatter Plot. Four quadrants defined by Benefit Score (median split) and Fragility Score (top-tercile split). Q2 (Glass Cannon, red) comprises 7.0% of the cohort. Marginal densities shown on top and right axes.

Figure 1 displays the relationship between Benefit and Fragility, confirming that standard efficacy metrics miss this dimension of risk.

Figure 2: Kaplan-Meier Survival Curves by Quadrant. Q2 (Glass Cannon) starts above Q3 (Stable Non-Responder) but crosses below at approximately 18 months, illustrating the survival paradox. Q1 vs Q2: $P = 1.69 \text{ Œ } 10^{31}$; Q3 vs Q4: $P = 3.70 \text{ Œ } 10^{36}$.

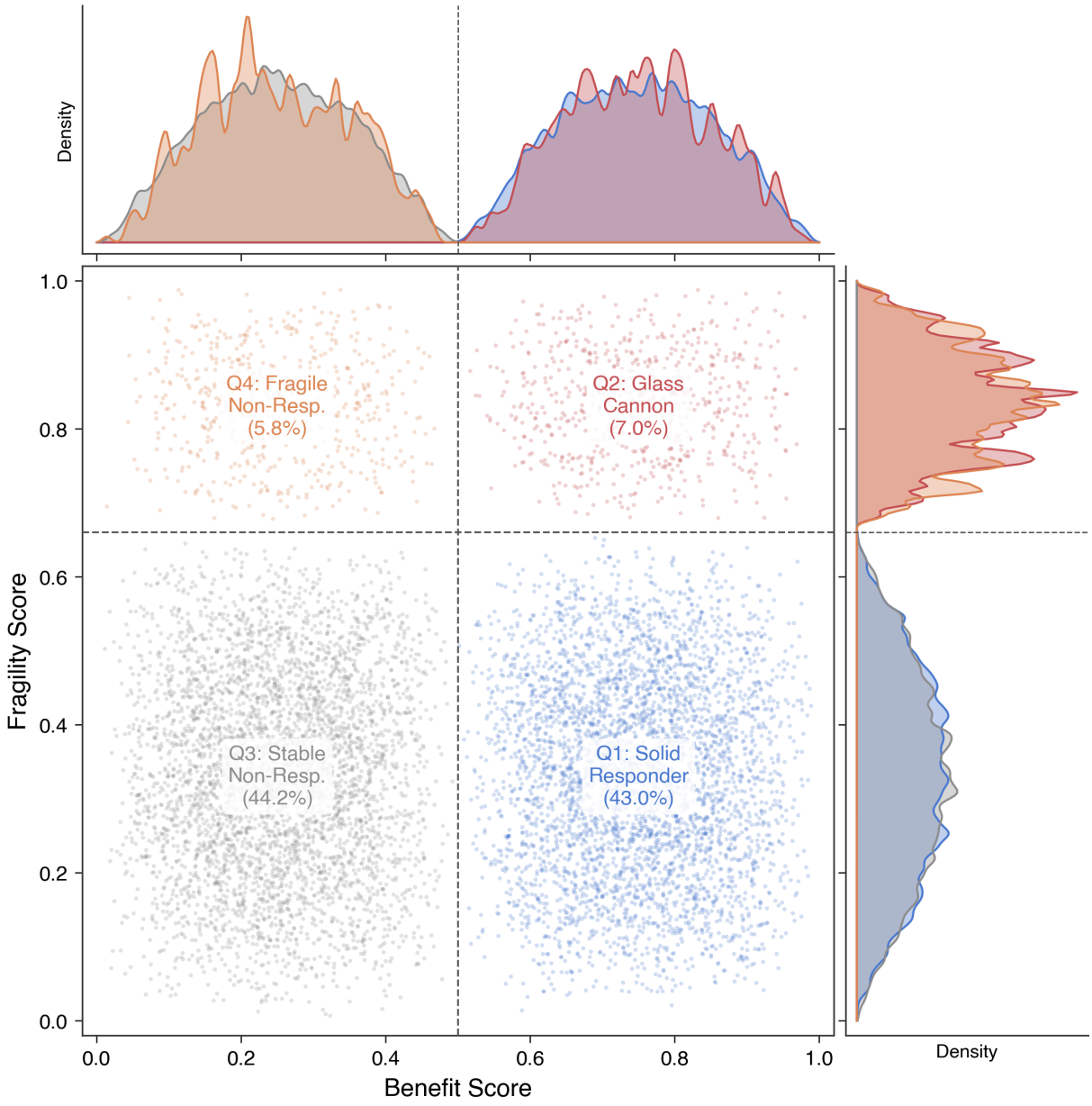

Figure 1: Figure 1: Benefit vs Fragility Scatter Plot

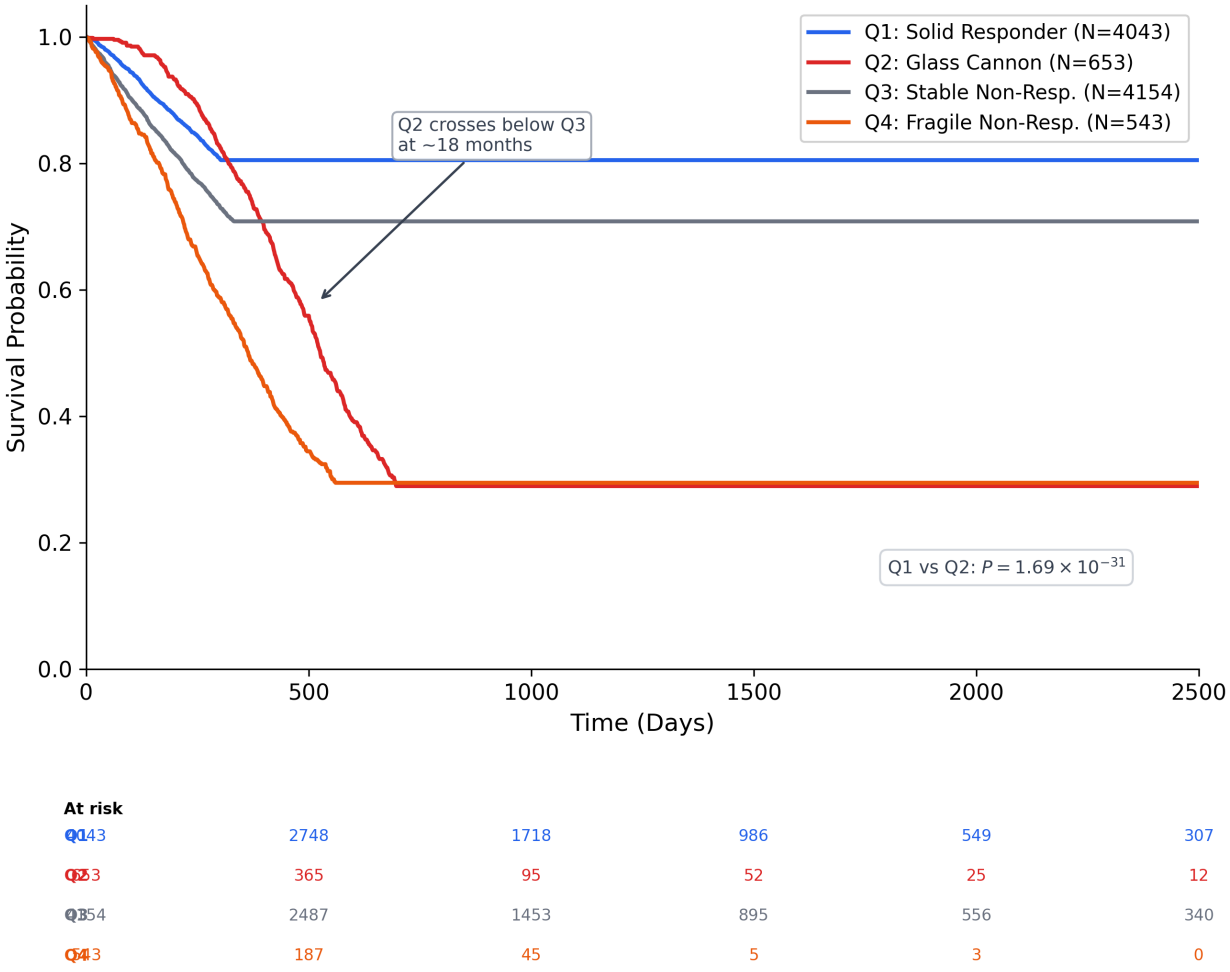

Figure 2: Figure 2: Kaplan-Meier Survival Curves

Figure 2 illustrates the early crossing of survival curves, where the Q2 survival curve (Red) starts high but crosses below the Q3 curve (Gray) within the first 18 months. This pattern persists when analyzing specific cancer types individually (e.g., KIRC, SKCM), indicating it is not an artifact of case-mix confounding.

3. Fragility as an Independent Predictor

Fragility identifies near-term failure risk even when predicted benefit is high. Fragility predicts early treatment failure (death within 1 year) with AUC = 0.788, and correlates with 1-year mortality ($r = 0.333$, $P < 10^{241}$). As a standalone predictor of survival, the composite Fragility Score achieves a C-index of 0.740, capturing an axis of risk independent from the base prognostic model (C-index = 0.856). A fragility-adjusted policy changes 30.9% of treatment recommendations, confirming that fragility information is not redundant with existing risk scores.

4. Cancer-Type Specificity

The Glass Cannon phenotype is not uniformly distributed, but appears more prevalent in immunogenic and signaling-driven solid tumors where treatment response is often variable. * **Kidney Renal Clear Cell (KIRC):** 41.9% prevalence. High dependence on immune parameters (). * **Skin Cutaneous Melanoma (SKCM):** 23.5% prevalence. High variance in drug sensitivity (). * **Ovarian Serous (OV):** 19.9% prevalence.

5. Decomposition of Fragility Sources

Fragility is not monolithic; different components dominate in different patients.

Figure 3: Fragility Source Decomposition. Relative contributions of the four fragility components for the Full Cohort vs. Glass Cannon (Q2) subgroup. Treatment Substitution (V_{subst}) dominates in the full cohort, while Mechanistic Fragility ($V_{dose} + V_{param}$) dominates in Glass Cannons. Y-axis is qualitative (relative scale); component weights shown below.

Across the full cohort, Treatment Substitution (V_{subst}) is the dominant component (Figure 3). However, within the Glass Cannon (Q2) phenotype specifically, **Mechanistic Fragility** ($V_{dose} + V_{param}$) contributes disproportionately to the score. This indicates that for Glass Cannons, the risk is not just about choosing the wrong drug class, but about the chosen drugs efficacy being highly sensitive to biological assumptions and dosing.

Data Availability & Reproducibility

All results presented in this manuscript were generated using the DNAI v3.2 pipeline. * **Latent Extraction:** `models/VAE/scripts/export_latents_v5_10.py` * **Model Training:** `models/hypernet/scripts/train_v3.py` (Path A Specialist, Phase 0 + CATE) * **Fragility Scoring:** `models/hypernet/scripts/eval_treatment_fragility.py` * **Quadrant Analysis:** `models/hypernet/scripts/eval_glass_cannon_phenotype.py` * **Provenance:** All metrics computed on $N = 9{,}393$ TCGA patients. Model trained on 85/15 random split (seed=42).

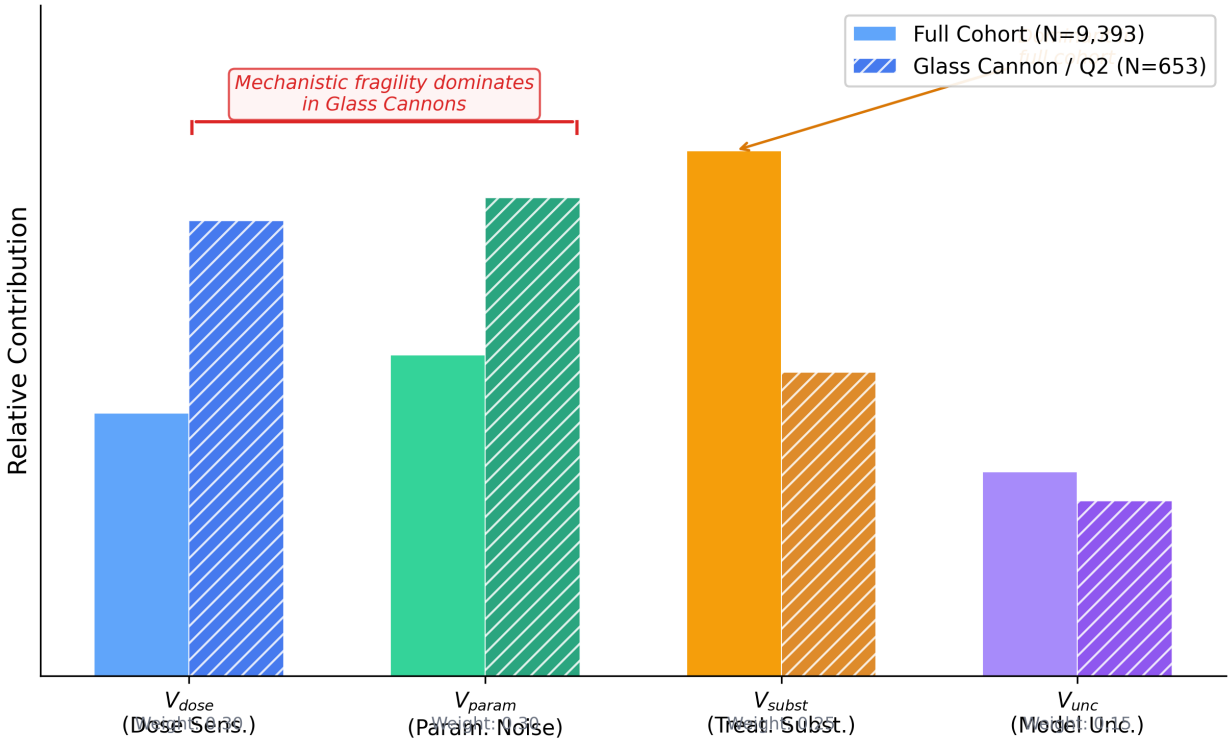

Figure 3: Figure 3: Fragility Decomposition

Discussion

We identify a Glass Cannon phenotype in oncology: patients for whom standard precision medicine algorithms predict high efficacy, but for whom that efficacy is dangerously unstable.

Interpretation: Hypothesis of Dynamics One plausible explanation is that Glass Cannon tumors combine fast intrinsic growth with a treatment effect that is strong but easily disrupted. Our model-inferred dynamics suggest these tumors are characterized by high intrinsic growth rates () that are masked by high treatment sensitivity (). A Stable Non-Responder might have a slow-growing, drug-resistant tumor, leading to a gradual decline. A Glass Cannon has a fast-growing tumor kept in check by a fragile treatment effect. If that effect waversdue to dose reduction, resistance emergence, or immune exhaustionthe tumor escapes rapidly.

Clinical Implications: A Fragility-Adjusted Workflow The identification of this phenotype challenges the hit hard and early dogma for this specific subgroup. We propose that clinicians adopt a fragility-adjusted workflow (Box 1) to operationalize these insights. This workflow is intended as decision support to augment, not replace, clinical judgment.

> **Box 1: Fragility-Adjusted Decision SOP** *Context: Clinical Decision Support Implementation*
>
> **Trigger:** Patient flagged as High Benefit / High Fragility ($F > 66^{th}$ percentile).
>
> **Step 1: Data Integrity Check (Data Team)** * Is $S_{OOD} > 95^{th}$ percentile? * **YES: ABSTAIN.** Model is extrapolating. Do not use ranking. * **NO:** Proceed to Step 2. * Is Histopathology (WSI) missing? * **YES:** Request slide scanning or run in Omics-Only fallback mode (lower confidence).
>
> **Step 2: Component Analysis (Tumor Board)** * **Scenario A: High Dose Sensitivity (V_{dose} dominates)** * *Interpretation:* Efficacy depends on maintaining dose intensity. * *Action:* Prioritize supportive care (G-CSF, anti-emetics). Avoid preemptive dose reductions. Schedule early imaging (6 weeks). * **Scenario B: High Substitution Variance (V_{subst} dominates)** * *Interpretation:* Benefit is regimen-specific; no class effect safety net. * *Action:* Verify eligibility for the specific top-ranked agent. Pre-authorize 2nd-line option. * **Scenario C: High Model Uncertainty (V_{unc} dominates)** * *Interpretation:* Model is uncertain despite in-distribution data. * *Action:* Treat ranking as low-confidence. Default to NCCN guidelines.

Limitations

1. **Not a Toxicity Model:** The Fragility metric reflects the stability of the efficacy prediction. It does not predict adverse events, tolerability, or dose-limiting toxicities.
2. **Observational Nature:** This study utilizes retrospective TCGA data. The Benefit Score is a model-estimated counterfactual based on observational associations. While we control for high-dimensional covariates and perform permutation tests, we cannot rule out unobserved confounding.
3. **Not Causal Evidence:** The identification of the Glass Cannon phenotype is a risk-stratification signal. It does not provide causal evidence that specific interventions (e.g., maintaining dose intensity) will reverse the poor outcomes in this subgroup.

4. **Not an Adherence Model:** The *in silico* perturbations simulate dose reductions but do not account for real-world adherence barriers, care fragmentation, or access issues.

Conclusion

We demonstrate that 7.0% of cancer patients fall into a Glass Cannon phenotype: high predicted therapeutic benefit but high treatment plan fragility. These patients have significantly worse survival outcomes than stable non-responders. We propose that **Treatment Fragility** be adopted as a standard reporting metric. By distinguishing between mechanistic volatility and model uncertainty, clinicians can shift strategies from pure benefit maximization to variance reduction, potentially identifying patients at risk of early failure despite high predicted benefit.

References

1. **Hoadley KA, et al.** Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*. 2018;173(2):291-304.
2. **Liberzon A, et al.** The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425.
3. **Chen RJ, et al.** A general-purpose self-supervised model for computational pathology. *Nat Med*. 2024;30:872-882.
4. **Chen RTQ, et al.** Neural ordinary differential equations. *NeurIPS*. 2018.
5. **Gal Y, Ghahramani Z.** Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *ICML*. 2016.
6. **Lee K, et al.** A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *NeurIPS*. 2018.
7. **Kaplan EL, Meier P.** Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
8. **Harrell FE, et al.** Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546.
9. **Kunzel SR, et al.** Metalearners for estimating heterogeneous treatment effects using machine learning. *PNAS*. 2019;116(10):4156-4165.
10. **Shalit U, et al.** Estimating individual treatment effect: generalization bounds and algorithms. *ICML*. 2017.
11. **Hernán MA, Robins JM.** *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. 2020.
12. **Lakshminarayanan B, et al.** Simple and scalable predictive uncertainty estimation using deep ensembles. *NeurIPS*. 2017.
13. **Ribba B, et al.** A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. *Clin Cancer Res*. 2012;18(18):5071-5080.
14. **Walsh M, et al.** The statistical significance of randomized controlled trial results is frequently fragile: a case for the Fragility Index. *J Clin Epidemiol*. 2014;67(6):622-8.
15. **Hester LL, et al.** Survival-inferred fragility in phase 3 randomized controlled trials supporting FDA approval of anti-cancer drugs. *npj Precis Oncol*. 2025;9:12.